SAFEGUARD

Harmonization of outcome extraction for heart failure across data sources in the SAFEGUARD project

Lorenza Scotti[1], Andrea Arfè[1], Antonella Zambon[1], Silvana Romio[1,2], Niklas Schmedt[3], Giorgia De Berardis[4], Cormac Sammon[5], Miguel Gil[6], Elisa Martin[6], Gema Requena[7], Corine de Vries[s5], John Seeger[8], Giampiero Mazzaglia[9], Irene Bezemer[10], Cristina Varas-Lorenzo[11], Mark Smits[12], Peter Rijnbeek[2], Miriam Sturkenboom[2], Giovanni Corrao[1]

[1] University Milano-Bicocca, Milan, Italy [2] Erasmus University Medical Center, Rotterdam, Netherlands ,[3] Leibniz-Institute for Prevention Research and Epidemiology – BIPS GmbH, Bremen, Germany , [4] Consorzio Mario Negri Sud, Santa Maria Imbaro, Italy, [5] University of Bath, Bath, United Kingdom, [6] Spanish Agency for Drugs and Medical Devices, Madrid, Spain, [7] University of Alcalá, Madrid, Spain, [8] The Brigham and Women's Hospital, Harvard Medical School, Boston, United States ,[9] Fondazione Scientifica SIMG-ONLUS, Firenze, Italy, [10] PHARMO Institute, Utrecht, Netherlands, [11] RTI Health Solutions, Barcelona, Spain,[12] VU University Medical Center, Amsterdam, Netherlands

Background

The Safety Evaluation of Adverse Reaction in Diabetes (SAFEGUARD) project aims to assess the relationship between use of glucose lowering agents and risk of selected events, including heart failure (HF), using data retrieved from different databases (DBs) from several EU countries and the USA. When combining data provided by several DBs, using different coding systems and considering different data collection approaches, an harmonization process is crucial to standardize the definition of the relevant events to ensure the direct comparability of the results.

Objectives

To harmonize the definition and the of HF in the different DBs involved in the SAFEGUARD project.

Methods

The data used to identify HF were retrieved from 9 DBs of different types (administrative, record linkage systems and general practitioner databases) from 5 EU countries and the USA. The study period ranged from 1999 to 2012. Each DB covered a different time period depending on its data availability and used different coding system to record the information. The main characteristics of the DBs involved in the project are reported in Table1.

Table 1 Main characteristics of the DBs involved in the SAFEGUARD project

Country	Database	DB Type*	Coding system	Study period	Target Population (milion)
Germany	GePaRD	Admin	ICD-10 GM	2004-2009	>14
Italy	HSD	GP	ICD-9 CM + Free Text	2000-2010	1.4
	LOMBARDY	Admin	ICD-9 CM	2000-2010	9
	PUGLIA	Admin	ICD-9 CM	2002-2009	5
Spain	BIFAP	GP	ICPC+ Free Text	2001-2009	3.2
The Netherlands	IPCI	GP	ICPC+ Free Text	2007-2012	1.1
	PHARMO	RL	ICD-9 CM	1999-2010	4
UK	CPRD	GP	READ code	2000-2011	8
USA	MEDICARE†	Admin	ICD-9 CM	2005-2008	>4

*Admin: Administrative DB; GP: General practitioner DB; RL: Record linkage system
†Data from MEDICARE is available only for patients older than 65 year due to the enrolment characteristic of the DB

The process leading to the harmonization of the HF definition began with the identification of the clinical definition of HF and the selection of all ICD-9, ICD-10, ICPC and READ codes and/or free-text terms suitable to identify the event of interest. Subsequently, the process proceeded through to the following two phases:

Extraction phase:

i. The DB specific codes used to identify HF were reviewed by the University of Milan-Bicocca with the support of other research partners.
ii. The data was extracted locally from each DB according to the reviewed codes and then processed using the software Jerboa©, developed by Erasmus MC.
iii. The Jerboa output provided the age and gender specific Incidence Rates (IRs) and standardized incidence rate (SIR) for HF per 100,000 person-years (PY), using the WHO world standard population as reference.

The source population was the entire population covered by the DBs. The IRs and SIRs were calculated without applying any restriction/exclusion criteria.

Harmonization phase:

This is an iterative procedure that comprise the following steps:

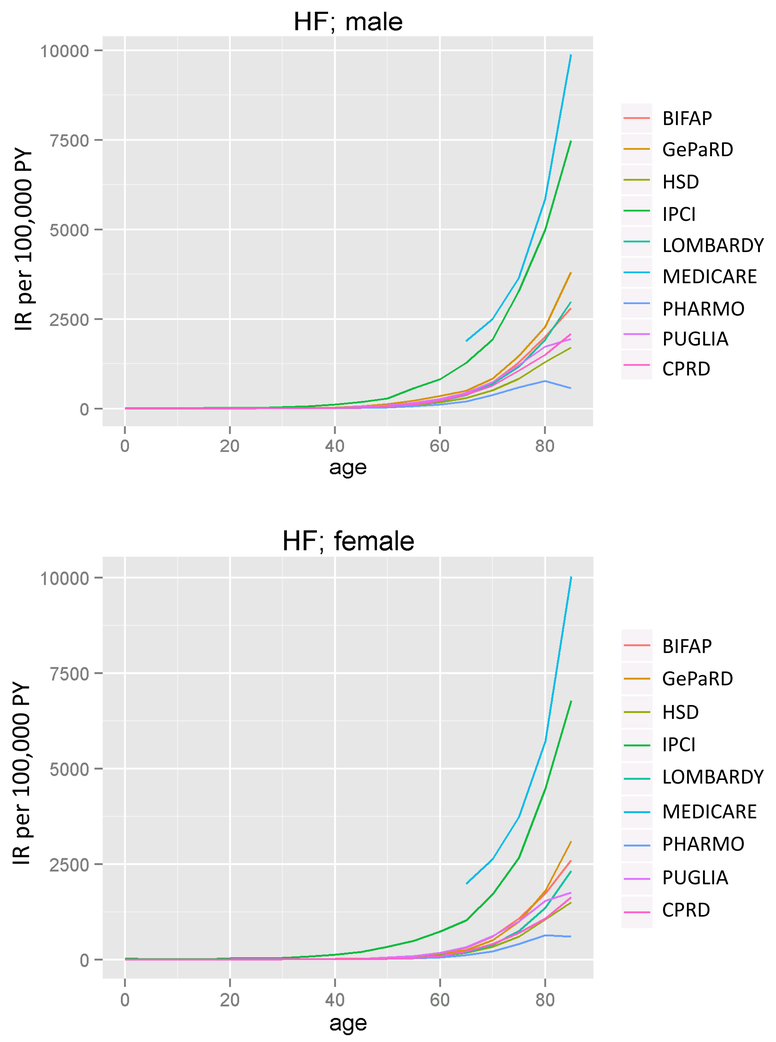
i. The HF code distribution was compared between DBs using the same coding system to highlight possible omissions or insertions of codes and extraction problems.
ii. Age and gender specific HF IRs were compared between databases and with those available in the scientific literature to evaluate their consistency.
iii. If changes in the extraction code were needed, they were implemented in the subsequent runs until no other modification were necessary.

All steps were repeated until an harmonized outcome definition was agreed upon.

Results

The source population of almost 47 million subjects accumulated 267,385,896 PY and generated 527,721 HF events during the study period.
The figure reports the age-specific HF IRs for all DBs, separately for males (upper panel) and females (lower panel).

Figure Age-specific HF IRs, by database.



- Males and females showed similar trends of HF IRs by age. For most of DBs, the IRs are of the order of 60 per 100,000 PY until the age of 50 years for males and 60 years for females and then start to increase with age. Excluding IPCI and MEDICARE, for the age class ≥85 years, the IRs range from 571 to 3818 per 100,000 PY for males and from 605 to 3104 per 100,000 PY for females.
- For what concern the comparison of the SIRs, except for IPCI (SIR= 312 per 100,000 PY) and MEDICARE (SIR= 3470 per 100,000 PY), the SIRs of HF ranged from 38 to 103 per 100,000 PY in Administrative DBs or RL system and from 66 to 99 per 100,000 PY in general practice DBs.

Conclusions

• The estimated IRs seem to be in line with those reported in the scientific literature[1,2], with the possible exception for those obtained from IPCI and Medicare. The higher IRs in Medicare and IPCI may be explained by misclassification.
• The SIRs obtained by Administrative, GP DBs and Record linkage system are consistent.
• Validation procedures will be applied in some DBs to verify the accuracy of the codes used in the definition of HF.

References

[1]Roger VL et al. Trends in heart failure incidence and survival in a community-based population. JAMA 2004:292:344-50
[2]Johansson, S et al., Incidence of Newly Diagnosed Heart failure in UK General Practice, Eur J Heart Failure 2001:3:225-31

Conflicts of interest

C. Varas-Lorenzo as employee of RTI Health Solutions participates in project advisory boards funded by pharmaceutical companies.
M. Sturkenboom is head of an academic unit that conducts research for pharmaceutical companies: Pfizer, Lilly, AstraZeneca, Boerhinger Ingelheim

Contact information

Lorenza Scotti, PhD
Department of Statistics and quantitative methods – University of Milano-Bicocca
Via Bicocca degli Arcimboldi, 8
21026, Milan, Italy
Phone: +39(0)264485846 mail: lorenza.scotti@statistica.unimib.it

Project website: **www.safeguard-diabetes.org**

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement nº 282521 - the SAFEGUARD project.